Anekant Education Society's

Tuljaram Chaturchand College of Arts, Science and Commerce, Baramati

(AUTONOMOUS)

Department of Statistics

Two Years Post Graduate Program

M.Sc. Data Science

Course Structure and Syllabus

Semester- I

Paper Code	Course Title	No. of Credits
PSDS111	Linear Algebra in Matlab	04
PSDS112	Probability Distributions	04
PSDS113	Optimization Techniques	04
PSDS114	Statistical Inference	04
PSDS115	Database Management System	04
PSDS116	Introduction to MATLAB and R	04

Semester- II

Paper Code	Course Title	No. of Credits
PSDS121	Design and Analysis of Experiments	04
PSDS122	Regression Analysis and Predictive Models	04
PSDS123	Statistical Quality Control	04
PSDS124	Computational Statistics	04
PSDS125	Bayesian Inference	04
PSDS126	Python and SQL Programming	04

Semester- III

Paper Code	Course Title	No. of Credits
PSDS231	Stochastic Models and Applications	04
PSDS232	Exploratory Multivariate Data Analysis	04
PSDS233	Time series analysis and Forecasting	04
PSDS234	Artificial Intelligence	04
PSDS235	Text Mining and Natural Language Processing	04
PSDS236	Data Visualization using Tableau	04

Semester- IV

Paper Code	Course Title	No. of Credits
PSDS241	Machine Learning	04
PSDS242	Discrete Data Analysis	04
PSDS243	Supply Chain & Logistics Analytics	04
PSDS244	Deep Learning	04
PSDS245	Thesis	08

PO1	**Disciplinary Knowledge:** Demonstrate comprehensive knowledge of the discipline that forms a part of a postgraduate programme. Execute strong theoretical and practical understanding generated from the specific programme in the area of work.
PO2	**Critical Thinking and Problem solving:** Exhibit the skill of critical thinking and understand scientific texts and place scientific statements and themes in contexts and also evaluate them in terms of generic conventions. Identify the problem by observing the situation closely, take actions and apply lateral thinking and analytical skills to design the solutions.
PO3	**Social competence:** Exhibit thoughts and ideas effectively in writing and orally; communicate with others using appropriate media, build effective interactive and presenting skills to meet global competencies. Elicit views of others, present complex information in a clear and concise way and help reach conclusions in group settings.
PO4	**Research-related skills and Scientific temper :** Infer scientific literature, build a sense of enquiry and able to formulate, test, analyse, interpret and establish hypothesis and research questions; and to identify and consult relevant sources to find answers. Plan and write a research paper/project while emphasizing on academics and research ethics, scientific conduct and creating awareness about intellectual property rights and issues of plagiarism.
PO5	**Trans-disciplinary knowledge:** Create new conceptual, theoretical and methodological understanding that integrates and transcends beyond discipline-specific approaches to address a common problem.
PO6	**Personal and professional competence:** Perform independently and also collaboratively as a part of a team to meet defined objectives and carry out work across interdisciplinary fields. Execute interpersonal relationships, self-motivation and adaptability skills and commit to professional ethics.
PO7	**Effective Citizenship and Ethics:** Demonstrate empathetic social concern and equity centred national development, and ability to act with an informed awareness of moral and ethical issues and commit to professional ethics and responsibility.
PO8	**Environment and Sustainability:** Understand the impact of the scientific solutions in societal and environmental contexts and demonstrate the knowledge of and need for sustainable development.
PO9	**Self-directed and Life-long learning:** Acquire the ability to engage in independent and life-long learning in the broadest context of socio-technological changes.

SYLLABUS (CBCS) FOR M.Sc. Data Science

(With effect from Academic Year 2022-2023)

Class : M. Sc. (Semester- I)
Paper Code : PSDS111
Paper : I
Title of Paper : Linear Algebra in Matlab
Credit : 4 credits
No. of lectures : 60

A) Course Objectives:

1. To understand basic matrix properties like rank, determinant, inverse and a special type of matrices
2. To introduce Gaussian / Gauss-Jordan elimination methods, LU factorization technique
3. To understand the concepts of vector space and subspaces.

B) Course Outcomes:

By the end of the course, students should be able to:

CO1 Having a clear understanding of the subject related concepts and contemporary issues.
CO2 Having computational thinking.
CO3 Ability to translate vast data into abstract concepts and to understand database reasoning.
CO4 Having problem-solving ability- solving social issues and engineering problems.
CO5 Find the matrix representation of a linear transformation given bases of the relevant vector spaces.
CO6 Compute inner products on a real vector space and compute angle and orthogonally in inner product spaces.
CO7 Understand the use of linear algebra and matrices in several important, modern applications of research and industrial problems involving statistics.

Unit 1

Vector: Vector addition, Scalar Vector multiplication, Inner Product, Complexity of Vector Computations Vector Spaces, Subspaces, Basis and dimension of a vector space, linear dependence and linear independence, spanning set. Linear transformation, kernel, range, Matrix Representation of a linear transformation, rank nullity theorem, change of basis and similar matrices. Inner-product spaces, orthogonal sets and bases, Orthogonal Projection, Gram-Schmidt orthogonalization process. Norms and Distance, Standard deviation, Angle, Complexity. **(15 L)**

Unit 2

Algebra of Matrices, Trace and Rank of a Matrix and their properties, Determinants, Inverse, Left and right inverses, Pseudo Inverse, nilpotent matrix , idempotent matrices and their properties, Eigen values and Eigen vectors, symmetric, orthogonal , , Gauss elimination, row

canonical form, diagonal form, triangular form, Gauss-Jordan-LU decomposition, solving systems of linear equations. **(15 L)**

Unit 3

L1 norm, L2 norm, regularization of norm, covariance matrix, Singular Vectors, Singular Value Decomposition, Best Rank k Approximations, algebraic and geometric multiplicities, Cayley-Hamilton theorem, Power Method for Singular Decomposition, Singular Vectors and Eigen Vectors, Applications of Singular Value Decomposition to Centering Data ,Principal Component Analysis, Ranking Documents and Web Pages, Discrete Optimization Problem.

(15 L)

Unit 4

Generalized inverses (g-inverses), Methods of constructing g-inverses, general solution to a system of linear equations. Sparse matrices, Linear Discriminant Analysis and Canonical Correlation Analysis Spectral decomposition, Quadratic forms, definiteness and related results with proofs. **(15 L)**

References Books:

1. Bapat, R.B. (2011).Linear Algebra and Linear Models. Springer and Hindustan Book Agency.
2. Stephen Boyd (Stanford University) and Lieven Vandenberghe (University of California, Los Angeles), Introduction to Applied Linear Algebra Vectors, Matrices and Least Squares , Cambridge University Press
3. John Chandler, Swarna Reddy, Algorithms for Data Science by Brian Steele, Springer International Publishing Switzerland 2016 .Kollo, T. and Rosen, D. von (2005). Advanced Multivariate Statistics with Matrices, Springer, and New York.
4. Kumaresan, S. (2000). Linear Algebra: A Geometric Approach, Prentice Hall
5. Lay, D. C. Lay, S. R. and Mc Donald, J. J. (2016) .Linear Algebra and Its Applications, Fifth Edition, Pearson, Boston.
6. Ramachandra Rao, A. and Bhimasankaram, P. (2000). Linear Algebra. Hindusta Book Agency
7. Rao, C. R. (1995). Linear Statistical Inference and Its Applications, Wiley
8. Searle, S. R. (1982). Matrix Algebra Useful for Statistics, John Wiley, New York.
9. G. Allaire and S. M. Kaber. Numerical Linear Algebra, Texts in Applied Mathematics, Springer, 2008. L. Hogben, Handbook of Linear Algebra, CRC Press/Taylor & Francis Group, 2014. Friedberg, S., Insel, A., and Spence, L., Linear Algebra, 5/e, Pearson, 2019.

Programme Outcomes and Course Outcomes Mapping:

Course Outcomes	Programme Outcomes (POs)								
	PO1	**PO2**	**PO3**	**PO4**	**PO5**	**PO6**	**PO7**	**PO8**	**PO9**
CO1	3	3			1	1		1	3
CO2		3			1	1		1	3
CO3		3			1	1		1	3
CO4		3	3		1	1	3	1	3
CO5		3			1	1		1	3
CO6		3			1	1		1	3
CO7		3		3	1	1		1	3

Weight: 1 - Partially related 2 - Moderately Related 3 - Strongly related

PO1. Disciplinary Knowledge

CO1. Having a clear understanding of the subject-related concepts and contemporary issues. (Weightage: 3 - Strongly Related)

Justification: Having a clear understanding of subject-related concepts contributes directly to disciplinary knowledge.

PO2. Critical Thinking and Problem Solving

All COs (Weightage: 3 - Strongly Related)

Justification: All outcomes involve critical thinking and problem-solving skills, from computational thinking to solving social and engineering problems.

PO3. Social Competence

CO4. Having problem-solving ability- solving social issues and engineering problems. (Weightage: 3 - Strongly Related)

Justification: Problem-solving ability, especially in the context of social issues, directly aligns with developing social competence.

PO4. Research-related Skills and Scientific Temper

CO7. Understand the use of linear algebra and matrices in several important, modern applications of research and industrial problems involving statistics. (Weightage: 3 - Strongly Related)

Justification: Understanding the use of linear algebra in research and industrial problems aligns closely with research-related skills and a scientific temper.

PO5. Trans-disciplinary Knowledge

All COs (Weightage: 1 - Partially Related)

Justification: The outcomes contribute more to technical competence and understanding within the discipline, with limited direct connections to trans-disciplinary knowledge.

PO6. Personal and Professional Competence

All COs (Weightage: 1 - Partially Related)

Justification: The outcomes contribute more to technical competence in the relevant subject matter, with limited direct connections to personal and professional aspects.

PO7. Effective Citizenship and Ethics

CO4. Having problem-solving ability- solving social issues and engineering problems. (Weightage: 3 - Strongly Related)

Justification: Solving social issues and engineering problems involves ethical considerations and aligns with effective citizenship.

PO8. Environment and Sustainability

All COs (Weightage: 1 - Partially Related)

Justification: The outcomes are more technical and less directly related to environmental or sustainability aspects.

PO9. Self-directed and Life-long Learning

All COs (Weightage: 3 - Strongly Related)

Justification: Continuous problem-solving, understanding of concepts, and engagement with contemporary issues align with self-directed and life-long learning.

SYLLABUS (CBCS) FOR M.Sc. Data Science

(2022 Pattern)

(With effect from Academic Year 2022-2023)

Class	: M. Sc. (Semester- I)
Paper Code	: PSDS112
Paper	: II
Title of Paper	: Probability Distributions
Credit	: 4 credits
No. of lectures	: 60

A) Course Objectives:

1. To incorporate the concepts of probability theory and its applications as the core material in building theoretical ideas along with the practical notion.

B) Course Outcomes:

Student will be able to:

CO1 Develop problem-solving techniques needed to calculate probability and conditional probability.

CO2 Formulate fundamental probability distribution and density functions, as well as functions of random variables, derive the probability density function of transformations.

CO3 Derive the expectation and conditional expectation, and describe their properties.

CO4 Integrate the intrinsic ideas of preliminary and advanced distributions to correlate with the real-world scenarios.

CO5 Understand various types of generating functions used in statistics.

CO6 Present the analysis of derived statistics to all audiences.

CO7 Develop problem-solving techniques needed to accurately calculate probabilities

TOPICS/CONTENTS:

Unit 1

Probability and Random variables

Introduction – Random Experiments, Empirical basis of probability, Algebra of events, laws of Probability; Conditional Probability, Independence, Bayes' law; Application of probability to business and economics. One-dimensional Random variable- Discrete and Continuous; Distribution functions and its properties; Bivariate Random Variables- Joint Probability

functions, marginal distributions, conditional distribution functions, Notion of Independence of Random variables **(15 L)**

Unit 2

Functions of Random Variables

Functions of random variables: introduction, distribution function technique, transformation technique: one variable, transformation technique: several variables, theory and applications.

Mathematical Expectation

Expectation, Variance, and Co-variance of random variables; Conditional expectation and conditional variance; Markov, Holder, Jensen and Chebyshev's Inequality, Weak Law of Large numbers, Strong law of large numbers and Kolmogorov theorem, Central Limit Theorem. **(15L)**

Unit 3

Generating Functions

Probability generating function (p.g.f.), moment generating function (m.g.f.), characteristic function (c.f.) Properties and Applications. Probability distributions of functions of random variables: one and two dimensions.

Sampling Distributions

Introduction, The sampling distribution of the Mean: Finite Populations, Sampling distribution of the proportion: Finite Populations, distribution of sample variance, the chi-square distribution, the t distribution, the F distribution, order statistics: properties, and applications, procedure of hypothesis testing. **(15L)**

Unit 4:

Discrete Distributions

Bernoulli, Binomial, Poisson, Geometric, Hypergeometric, Negative Binomial, Multinomial, distributions and Discrete Uniform distribution - definition, properties and applications with numerical problems.

Continuous Distributions

Uniform, Normal distribution function, Exponential, Gamma, Beta distributions (First and Second kind), Weibull, Cauchy and Laplace distributions, lognormal, logistic, Pareto, Chi-sqaure and Rayleigh distribution functions - definition, properties and applications; concept of truncated distributions. **(15L)**

References Books:

1. Parimal Mukhopadhyay; An Introduction to the Theory of Probability, World scientific, 2012.

2. Irwin Miller, Marylees Miller, John E. Freund's; Mathematical Statistics, Pearson, 2017.
3. Fetsje Bijma, Marianne Jonker and Aadvander Vaart; Introduction to Mathematical Statistics, Amsterdam University Press, 2018.
4. Krishnamoorthy, K., Handbook of Statistical Distributions with Applications, Chapman & Hall/CRC, 2006.
5. Shanmugam, R., Chattamvelli, R. Statistics for scientists and engineers, John Wiley, 2015.

Programme Outcomes and Course Outcomes Mapping:

Course Outcomes	Programme Outcomes (POs)								
	PO1	**PO2**	**PO3**	**PO4**	**PO5**	**PO6**	**PO7**	**PO8**	**PO9**
CO1	3		1	1	1	1	1	1	3
CO2		3	1	1	1	1	1	1	3
CO3			1	1	1	1	1	1	3
CO4			1	1	1	1	1	1	3
CO5			1	1	1	1	1	1	3
CO6			1	1	1	1	1	1	3
CO7			1	1	1	1	1	1	3

Weight: 1 - Partially related 2 - Moderately Related 3 - Strongly related

PO1. Disciplinary Knowledge

CO1. Having a clear understanding of the subject-related concepts and contemporary issues. (Weightage: 3 - Strongly Related)

Justification: Having a clear understanding of subject-related concepts contributes directly to disciplinary knowledge.

PO2. Critical Thinking and Problem Solving

CO2. Develop problem-solving techniques needed to calculate probability and conditional probability. (Weightage: 3 - Strongly Related)

Justification: Developing problem-solving techniques in probability directly aligns with critical thinking and problem-solving skills.

PO3. Social Competence

All COs (Weightage: 1 - Partially Related)

Justification: The outcomes are more technical and focused on probability concepts, with less direct relevance to social competence.

PO4. Research-related Skills and Scientific Temper

All COs (Weightage: 1 - Partially Related)

Justification: The outcomes are more focused on theoretical aspects of probability rather than practical research-related skills.

PO5. Trans-disciplinary Knowledge

All COs (Weightage: 1 - Partially Related)

Justification: The outcomes contribute more to technical competence within the discipline, with limited direct connections to trans-disciplinary knowledge.

PO6. Personal and Professional Competence

All COs (Weightage: 1 - Partially Related)

Justification: The outcomes contribute more to technical competence in probability concepts, with limited direct connections to personal and professional aspects.

PO7. Effective Citizenship and Ethics

All COs (Weightage: 1 - Partially Related)

Justification: The content is more technical and less directly related to effective citizenship and ethics.

PO8. Environment and Sustainability

All COs (Weightage: 1 - Partially Related)

Justification: The outcomes are more focused on probability concepts than on environmental or sustainability aspects.

PO9. Self-directed and Life-long Learning

All COs (Weightage: 3 - Strongly Related)

Justification: Continuous problem-solving, understanding of probability concepts, and engagement with contemporary issues align with self-directed and life-long learning.

SYLLABUS (CBCS) FOR M.Sc. Data Science

(With effect from Academic Year 2022-2023)

Class : M. Sc. (Semester- I)
Paper Code : PSDS113
Paper : III
Title of Paper : Optimization Techniques
Credit : 4 credits
No. of lectures : 60

A) Course Objectives:

1. To familiarize the students with some basic concepts of optimization techniques and approaches.
2. To formulate a real-world problem as a mathematical programming model.

B) Course Outcomes:

Students are able to

CO1 Identify and develop operational research models from the verbal description of the real system.

CO2 Understand the characteristics of different types of decision-making environments and decision-making approaches.

CO3 Apply optimization techniques to take correct decision.

CO4 Formulate a real-world problem as a mathematical programming model.

CO5 Develop the model formulation and applications are used in solving decision problems.

CO6 Solve specialized linear programming problems like the transportation and assignment Problems.

CO7 Apply optimization techniques to real world problem.

TOPICS/CONTENTS:

Unit 1

Introduction to Operations Research

Introduction to Mathematical models of Operation Research, Scope and applications of Operation Research, Phases of Operation Research, Characteristics of Operation Research, and Limitations of Operation Research**.**

Linear Programming

Introduction, Properties of Linear Programming, Basic assumptions, Mathematical formulation of Linear Programming, Limitations or constraints, Methods for the solution of LP Problem, Graphical Method, Simplex Method, Big M Method. **(15)**

Unit 2

Dual Linear Programming

Introduction to Primal and Dual problem, Dual problem properties, Solution techniques of Dual problem, Dual Simplex method, Relations between direct and dual problem, Economic interpretation of Duality. **(15L)**

Unit 3

Transportation and Assignment Models

Introduction to transportation problem (TP), Balanced and Unbalanced TP, Methods of basic feasible solution, Optimal solution, MODI method. Assignment problem, Hungarian Method.

Network Analysis

Basic concepts, Construction of Network, Rules and precautions, CPM and PERT Networks, obtaining of critical path, probability and cost consideration, advantages of Network. **(15L)**

Unit 4

Theory of Games

Introduction and terminology of Two Person Zero-Sum Game, Solution of games with saddle points and without saddle points, 2×2 games, dominance principle, m×2 and 2×n games, Graphical method.

Hyper parameter optimization

Gradient of a function, Steepest descent method, Nelder Mead's Simplex search method, Newtons method. **(15L)**

References Books:

1. S.D. Sharma (2000), Operations Research, Nath& Co., Meerut.MauriceSolient, Arthur Yaspen, Lawrence Fridman, (2003), OR methods and Problems,New Age International Edition.
2. J K Sharma (2007), Operations Research Theory & Applications, 3e, Macmillan India Ltd. P. SankaraIyer, (2008), Operations Research, Tata McGraw-Hill.
3. Taha, H.A., Operations Research: An Introduction, Prentice Hall of India, 9th Edition,2010.
4. A Ravindran, Don T Philips and James J Solberg, Operations Research: Principles and Practice, 2nd edition, John Wiley and sons, 2007.
5. L.S.Srinath, PERT and CPM Principles and Applications, Affiliated East-West Press(Pvt.) Ltd, 3rd edition, 2001

Programme Outcomes and Course Outcomes Mapping:

Course Outcomes	Programme Outcomes (POs)								
	PO1	**PO2**	**PO3**	**PO4**	**PO5**	**PO6**	**PO7**	**PO8**	**PO9**
CO1	3		1	1	1	1	1	1	3
CO2		3	1	1	1	1	1	1	3
CO3			1	1	1	1	1	1	3
CO4			1	1	1	1	1	1	3
CO5			1	1	1	1	1	1	3
CO6			1	1	1	1	1	1	3
CO7			1	1	1	1	1	1	3

Weight: 1 - Partially related 2 - Moderately Related 3 - Strongly related

PO1. Disciplinary Knowledge

CO1. Identify and develop operational research models from the verbal description of the real system. (Weightage: 3 - Strongly Related)

Justification: Identifying and developing operational research models directly contributes to disciplinary knowledge.

PO2. Critical Thinking and Problem Solving

CO2. Understand the characteristics of different types of decision-making environments and decision-making approaches. (Weightage: 3 - Strongly Related)

Justification: Understanding decision-making environments and approaches is integral to critical thinking and problem-solving skills.

PO3. Social Competence

All COs (Weightage: 1 - Partially Related)

Justification: The outcomes are more technical and focused on optimization and decision-making concepts, with less direct relevance to social competence.

PO4. Research-related Skills and Scientific Temper

All COs (Weightage: 1 - Partially Related)

Justification: The outcomes are more focused on applying optimization techniques and decision-making rather than on practical research-related skills.

PO5. Trans-disciplinary Knowledge

All COs (Weightage: 1 - Partially Related)

Justification: The outcomes contribute more to technical competence within the discipline, with limited direct connections to trans-disciplinary knowledge.

PO6. Personal and Professional Competence

All COs (Weightage: 1 - Partially Related)

Justification: The outcomes contribute more to technical competence in operational research and decision-making, with limited direct connections to personal and professional aspects.

PO7. Effective Citizenship and Ethics

All COs (Weightage: 1 - Partially Related)

Justification: The content is more technical and less directly related to effective citizenship and ethics.

PO8. Environment and Sustainability

All COs (Weightage: 1 - Partially Related)

Justification: The outcomes are more focused on optimization and decision-making concepts than on environmental or sustainability aspects.

PO9. Self-directed and Life-long Learning

All COs (Weightage: 3 - Strongly Related)

Justification: Continuous problem-solving, understanding decision-making approaches, and applying optimization techniques align with self-directed and life-long learning.

SYLLABUS (CBCS) FOR M.Sc. Data Science
(2022 Pattern)
(With effect from Academic Year 2022-2023)

Class : M. Sc. (Semester- I)
Paper Code : PSDS114
Paper : IV
Title of Paper : Statistical Inference
Credit : 4 credits
No. of lectures : 60

A) Course Objectives:

Students successfully completing this course will be able to:

1. Get basic understanding about statistical models and their use.
2. Apply linear and regression models depending upon the problem context.

B) Course Outcomes:

At the end of the course students will be able to:

CO1 Learn the approaches to point estimation of parameters.

CO2 Understand the concept of interval estimation and confidence intervals.

CO3 Basic concepts in tests of hypotheses.

CO4 Get a better understanding of probabilistic models.

CO5 Implement different tree based models.

CO6 Derive inference from different statistical data sets.

CO7 Develop a strong theoretical understanding of statistical concepts such as sufficiency, estimation, hypothesis testing, confidence intervals, and Bayesian estimation.

Unit 1

Introduction

Population, sample, parameter and statistic; characteristics of a good estimator, Unbiasedness, Sufficiency – Factorization Theorem – Minimal sufficiency, Efficiency – Most Efficient estimator, likelihood equivalence, applications of Lehmann-Scheffe's Theorem, Rao-Blackwell Theorem and, Consistency –Invariance property of Consistent estimator, sufficient condition for consistency uniformly minimum variance unbiased Estimator. **(15 L)**

Unit 2

Point Estimation

Point Estimation- Estimator, Estimate, Methods of point estimation – Maximum likelihood Method (the asymptotic properties of ML estimators are not included), and large sample Properties of ML estimator (without proof) -applications, Method of moments, method of

Least squares, method of minimum chi-square and modified minimum chi-square asymptotic, Maximum Likelihood Estimation and applications. **(15 L)**

Unit 3

Testing of hypotheses

Types of errors, power of a test, most powerful tests, Neyman–Pearson Fundamental Lemma and its applications; Notion of Uniformly most powerful tests; Likelihood Ratio tests, Description and property of LR tests - Application to standard distributions.

Large sample tests

Large sample properties, Tests of significance (under normality assumption) Test for a Population mean, proportion, Test for equality of two means, proportions, Test for Variance, Test for correlation, Test for Regression.

Small sample tests

Student's t-test, test for a population mean, equality of two population means, paired t-test, F-test For equality of two population variances, Chi-square test for goodness of Fit and test for independence of attributes, χ^2 test for testing variance of a Normal Distribution Analysis of Variance.

Unit 4

Interval estimation confidence limits and confidence coefficient;

Duality between acceptance region of a test and a confidence interval, Construction of confidence intervals for population proportion (Small and large samples) and between two population proportions (large samples), Confidence intervals for mean and variance of a normal population; Difference between the Mean and ratio of two normal populations.

Non-parametric tests Sign test, Signed rank test, Median test, Mann-Whitney test, Run test and one sample Kolmogorov – Smirnov test, Kruskal – Wallis H test (Description, properties and applications only). **(15 L)**

All topics to be covered using R software. Manual calculations are not expected.

Reference Books:

1. Casella G. and Beregar R.L. (2002) Statistical Inference, 2nd Edition (Duxbury Advanced Series)
2. Dudewitz E.J. & Mishra S.N.(1988) Modern Mathematical Statistics (John Wiley)
3. Kale B.K. (1999) A First course on Parametric Inference (Narosa)
4. Lehman E.L (1988) Theory of point estimation (John Wiley)
5. Lehman E.L(1986) Testing of Statistical hypotheses (John Wiley)
6. Rohatagi V.K. (1976) Introduction to theory of probability & mathematical statistics (John Wiley & sons)

7. Dasgupta A. (2008), Asymptotic Theory of Statistics and Probability, Spinger-Verlag, New York.

Programme Outcomes and Course Outcomes Mapping:

Course Outcomes	Programme Outcomes (POs)								
	PO1	**PO2**	**PO3**	**PO4**	**PO5**	**PO6**	**PO7**	**PO8**	**PO9**
CO1	3	3		3	1	1	1	1	3
CO2	3	3		3	1	1	1	1	3
CO3	3	3		3	1	1	1	1	3
CO4	3	3		3	1	1	1	1	3
CO5	3	3		3	1	1	1	1	3
CO6	3	3		3	1	1	1	1	3
CO7	3	3	2	3	1	1	1	1	3

Weight: 1 - Partially related 2 - Moderately Related 3 - Strongly related

PO1. Disciplinary Knowledge

CO1. Learn the approaches to point estimation of parameters. (Weightage: 3 - Strongly Related)

Justification: Learning approaches to point estimation is fundamental to building disciplinary knowledge in statistics.

PO2. Critical Thinking and Problem Solving

All COs (Weightage: 3 - Strongly Related)

Justification: All outcomes involve critical thinking and problem-solving skills, from understanding statistical concepts to implementing tree-based models.

PO3. Social Competence

All COs (Weightage: 1 - Partially Related)

Justification: The outcomes are more technical and focused on statistical methodologies, with less direct relevance to social competence.

PO4. Research-related Skills and Scientific Temper

All COs (Weightage: 3 - Strongly Related)

Justification: The outcomes align closely with research-related skills and the development of a scientific temper in the context of statistical concepts and modeling.

PO5. Trans-disciplinary Knowledge

CO5. Implement different tree-based models. (Weightage: 2 - Moderately Related)

Justification: Implementing tree-based models involves trans-disciplinary aspects as it applies statistical concepts to machine learning approaches.

PO6. Personal and Professional Competence

All COs (Weightage: 1 - Partially Related)

Justification: The outcomes contribute more to technical competence in statistics and data analysis, with limited direct connections to personal and professional aspects.

PO7. Effective Citizenship and Ethics

All COs (Weightage: 1 - Partially Related)

Justification: The content is more technical and less directly related to effective citizenship and ethics.

PO8. Environment and Sustainability

All COs (Weightage: 1 - Partially Related)

Justification: The outcomes are more focused on statistical methodologies and data analysis than on environmental or sustainability aspects.

PO9. Self-directed and Life-long Learning

All COs (Weightage: 3 - Strongly Related)

Justification: Understanding statistical concepts and methodologies requires ongoing self-directed learning and adaptation.

SYLLABUS (CBCS) FOR M.Sc. Data Science
(2022 Pattern)
(With effect from Academic Year 2022-2023)

Class : M. Sc. (Semester- I)
Paper Code : PSDS115
Paper : V
Title of Paper : Database Management System
Credit : 4 credits
No. of lectures : 60

A) Course Objectives:

Students successfully completing this course will be able to:

1. Recognize database system design and implementation.
2. Examine the logical, physical, and database modelling designs..

B) Course Outcomes:

By the end of the course, students should be able to:

1. Be familiar with the fundamentals of database concepts and database management systems.
2. Utilize conceptual modelling techniques, like as the ER model and relational model, to model the data requirements for an application.
3. Write SQL commands to create tables, insert, update, delete and querying data.
4. Developing an understanding of key DBMS principles.
5. Recognize how to create, modify, and query databases for data.
6. Proficiency in designing conceptual database models using the Entity-Relationship (E-R) model.
7. Apply constraints like key constraints and mapping constraints in ER modeling.

Unit 1

Introduction to File organization &DBMS, Database-system Applications, Purpose of Database Systems, Types of file Organization, File system Vs. DBMS, Data models, Levels of abstraction, Data in dependence, Structure of DBMS, Users of DBMS, Database Architecture, Speciality Databases.

(15L)

Unit 2

Structure of Relational Databases, Database Schema, Keys, Relational Operations, Conceptual Design (E-R model), Overview of DB design, ER data model (entities, attributes, entity sets, relations, relationship sets),Additional constraints (Key constraints, Mapping constraints), Conceptual design using ER modelling. Relational data model, Conversion of ER to Relational model, Integrity constraints, Relational algebra, Preliminaries. **(15L)**

Unit 3

Introduction to SQL, Basic structure, Set operations, Aggregate functions, Null values , PL/PgSqL: Data types, Language structure, Operations with SQL, Nested Sub queries, Modifications to Database, DDL and DML commands with examples, SQL mechanisms for joining. **(15L)**

Unit 4

Intermediate and advanced SQL: Join Expressions- Join conditions, Outer joins, Join types and conditions, Views- View definition, using views in SQL queries, Materialized views, update a view 4.3 Create table extensions, Schemas, Catalogs and Environments, The relational Algebra, The tuple relational calculus. **(15L)**

Reference Books:

1. Abraham Silberschatz, Henry F. Korth, S. Sudarashan, Database System Concepts, McGraw-Hill International Edition, Sixth Edition
2. Elmasri, Navathe, Fundamentals of Database Systems, Pearson Education, Third Education
3. Ramakrishnan, Gehrke, Database Management Systems, McGraw Hill International Edition, Third Edition
4. Peter Rob, Carlos Coronel, Database System Concepts, Cengage Learning, India Edition
5. S.K. Singh, "Database Systems Concepts, Design and Applications", First Edition, Pearson Education, 2006
6. Redmond, E. & Wilson, Seven Databases in Seven Weeks: A Guide to Modern Databases and the No SQL Movement Edition:1st Edition.
7. Shamkant B. Navathe, Ramez Elmasri,(2010), Database Systems, ISBN:9780132144988, PEARSON HIGHER EDUCATION
8. Richard Stones, Neil Matthew, (2005), Beginning Databases with PostgreSQL: From Novice to Professional, ISBN:9781590594780, Apress
9. Korry, Douglas, (2005), Postgre SQL, ISBN:9780672327568, Sams Publishing.
10. Joshua D. Drake, John C. Worsley, Practical Postgre SQL, (2002), ISBN:9788173663925 O'Reilly Media, Inc., ISBN: 9781565928466.

Programme Outcomes and Course Outcomes Mapping:

Course Outcomes	Programme Outcomes (POs)								
	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9
CO1	3	3		3	1	1	1	1	3
CO2	3	3		3	1	1	1	1	3
CO3	3	3		3	1	1	1	1	3
CO4	3	3		3	1	1	1	1	3
CO5	3	3		3	1	1	1	1	3
CO6	3	3		3	1	1	1	1	3
CO7	3	3	2	3	1	1	1	1	3

Weight: 1 - Partially related 2 - Moderately Related 3 - Strongly related

PO1. Disciplinary Knowledge

CO1. Be familiar with the fundamentals of database concepts and database management systems. (Weightage: 3 - Strongly Related)

Justification: Being familiar with the fundamentals of database concepts and management systems is fundamental to building disciplinary knowledge in database management.

PO2. Critical Thinking and Problem Solving

CO2. Utilize conceptual modeling techniques, like the ER model and relational model, to model the data requirements for an application. (Weightage: 3 - Strongly Related)

Justification: Utilizing conceptual modeling techniques requires critical thinking to model data effectively.

PO3. Social Competence

All COs (Weightage: 1 - Partially Related)

Justification: The outcomes are more technical and focused on database management, with less direct relevance to social competence.

PO4. Research-related Skills and Scientific Temper

All COs (Weightage: 3 - Strongly Related)

Justification: The outcomes align closely with research-related skills and the development of a scientific temper in the context of database management.

PO5. Trans-disciplinary Knowledge

All COs (Weightage: 1 - Partially Related)

Justification: The outcomes are more specific to database management and have limited trans-disciplinary connections.

PO6. Personal and Professional Competence

All COs (Weightage: 1 - Partially Related)

Justification: The outcomes contribute more to technical competence in database management, with limited direct connections to personal and professional aspects.

PO7. Effective Citizenship and Ethics

All COs (Weightage: 1 - Partially Related)

Justification: The content is more technical and less directly related to effective citizenship and ethics.

PO8. Environment and Sustainability

All COs (Weightage: 1 - Partially Related)

Justification: The outcomes are more focused on database management and less on environmental or sustainability aspects.

PO9. Self-directed and Life-long Learning

All COs (Weightage: 3 - Strongly Related)

Justification: Understanding database management requires ongoing self-directed learning and adaptation.

SYLLABUS (CBCS) FOR M.Sc. Data Science

(2022 Pattern)

(With effect from Academic Year 2022-2023)

Class : M. Sc. (Semester- I)
Paper Code : PSDS116
Paper : VI
Title of Paper : Introduction to MATLAB and R
Credit : 4 credits
No. of lectures : 60

A) Course Objectives:

1. To gain knowledge about MATLAB and R software for linear algebra, probability distribution and statistical inference.

B) Course Outcomes:

CO1 Students will be able to solve linear algebra problems using MATLAB/R - software.

CO2 Students will be able to draw model sample from distributions.

CO3 Students will understand asymptotic behaviour of the estimators, find and verify the consistent estimator and consistency.

CO4 Understanding how to compute determinants and ranks of matrices using partitioning techniques.

CO5 Apply partitioning methods to efficiently calculate determinants and ranks of higher-order matrices.

CO6 calculate determinants and ranks of matrices using partitioning methods for matrices of higher order.

CO7 Understanding the significance of determinants and ranks in linear algebra and their applications in solving systems of equations.

Sr. no.	Title of Experiments
1.	Calculation of determinant and rank of higher order by partitioning method
2.	Calculation of inverse of matrices of higher order.
3.	Calculation of Moore-Penrose inverse and g-inverse.
4.	Solution of simultaneous equations.
5.	Eigen value, Eigen vectors, Spectral decomposition, Power of matrix
6.	Plotting likelihood function and obtaining MLE.
7.	Plot density function, Distribution function and computation of probability of events related to standard probability distribution.

8.	Model sampling from Gamma, Chi square, Weibull, Lognormal probability distribution
9.	Model sampling from Discrete and continuous distribution.
10.	Verification of Consistency and CAN estimator.
11.	Comparison of consistent estimator based on MSE and sample size.
12.	Power function of large sample test. (LR, Wald, Rao)

Programme Outcomes and Course Outcomes Mapping:

Course Outcomes	Programme Outcomes (POs)								
	PO1	**PO2**	**PO3**	**PO4**	**PO5**	**PO6**	**PO7**	**PO8**	**PO9**
CO1	3	3	1	3	1	1	1	1	3
CO2		3	1	3	1	1	1	1	3
CO3		3	1	3	1	1	1	1	3
CO4		3	1	3	1	1	1	1	3
CO5		3	1	3	1	1	1	1	3
CO6		3	1	3	1	1	1	1	3
CO7		3	1	3	1	1	1	1	3

Weight: 1 - Partially related 2 - Moderately Related 3 - Strongly related

PO1. Disciplinary Knowledge

CO1. Students will be able to solve linear algebra problems using MATLAB/R - software. (Weightage: 3 - Strongly Related)

Justification: Proficiency in using MATLAB/R for linear algebra problems directly enhances disciplinary knowledge in mathematics and computational tools.

PO2. Critical Thinking and Problem Solving

All COs (Weightage: 3 - Strongly Related)

Justification: All outcomes involve critical thinking and problem-solving skills, especially in the context of linear algebra and statistical estimation.

PO3. Social Competence

All COs (Weightage: 1 - Partially Related)

Justification: The outcomes are more technical and focused on mathematical and statistical concepts, with limited direct relevance to social competence.

PO4. Research-related Skills and Scientific Temper

All COs (Weightage: 3 - Strongly Related)

Justification: The outcomes align closely with research-related skills and the development of a scientific temper in the context of statistical estimation and linear algebra.

PO5. Trans-disciplinary Knowledge

All COs (Weightage: 1 - Partially Related)

Justification: The outcomes are specific to mathematical and statistical concepts, and there is limited trans-disciplinary knowledge incorporated.

PO6. Personal and Professional Competence

All COs (Weightage: 1 - Partially Related)

Justification: The outcomes contribute more to technical competence in linear algebra and statistical estimation, with limited direct connections to personal and professional aspects.

PO7. Effective Citizenship and Ethics

All COs (Weightage: 1 - Partially Related)

Justification: The content is more technical and less directly related to effective citizenship and ethics.

PO8. Environment and Sustainability

All COs (Weightage: 1 - Partially Related)

Justification: The outcomes are more focused on linear algebra and statistical concepts and less on environmental or sustainability aspects.

PO9. Self-directed and Life-long Learning

All COs (Weightage: 3 - Strongly Related)

Justification: Mastering linear algebra and statistical estimation requires ongoing self-directed learning and adaptation, aligning well with the lifelong learning goal.